

CMDM-VAC: IMPROVING A PERCEPTUAL QUALITY METRIC FOR 3D GRAPHICS BY INTEGRATING A VISUAL ATTENTION COMPLEXITY MEASURE

Yana Nehmé[†] Mona Abid^{*} Guillaume Lavoué[†] Matthieu Perreira Da Silva^{*} Patrick Le Callet^{*}

[†] LIRIS CNRS, University of Lyon, France

^{*} LS2N CNRS, University of Nantes, France

ABSTRACT

Many objective quality metrics have been proposed over the years to automate the task of subjective quality assessment. However, few of them are designed for 3D graphical contents with appearance attributes; existing ones are based on geometry and color measures, yet they ignore the visual saliency of the objects. In this paper, we combined an optimal subset of geometry-based and color-based features, provided by a state-of-the-art quality metric for 3D colored meshes, with a visual attention complexity feature adapted to 3D graphics. The performance of our proposed new metric is evaluated on a dataset of 80 meshes with diffuse colors, generated from 5 source models corrupted by commonly used geometry and color distortions. With our proposed metric, we showed that the use of the attentional complexity feature brings a significant gain in performance and better stability.

Index Terms— Perceptual quality metric, 3D graphics, saliency, visual attention complexity, diffuse color.

1. INTRODUCTION

Nowadays, 3D Graphical Contents (3DGCs) are widely deployed in several domains such as digital entertainment, architecture, medicine, etc. These data undergo diverse processing operations, which introduce distortions that may impact the visual quality of the final rendering and thus the user’s Quality of Experience (QoE). In this sense, it is fundamental to evaluate the visual quality of 3DGCs as perceived by human observers. While subjective quality assessment tests can be very time-consuming and tedious, many objective quality metrics have been proposed over the years that automate the task. Most of these metrics were designed for 3D contents without appearance attributes; they rely on its geometry characteristics [1]. With respect to 3D content with color or material information, very few works have been published: metrics proposed for meshes with diffuse texture [2, 3] are usually based on a combination of mesh quality and texture quality. Very recently, the Color Mesh Distortion

Measure (CMDM) [4] was proposed for 3D meshes with diffuse colors. It is the first metric that operates entirely on the mesh domain. This metric incorporates perceptually relevant curvature-based and color-based features.

As can be seen, the mentioned model-based metrics (i.e., they operate on the 3D model itself and its attributes) are based on geometry and color/texture attributes. They ignore the visual saliency of the objects. In fact, visual saliency is an important feature of the human visual system. It describes the human attention distribution or the eye movements for a given scene [5]. We can use the visual saliency to determine perceptually salient regions. Finding these perceptually important regions has become a useful tool for many applications such as view-point selection [6], mesh simplification [7], segmentation [8], etc. In a recent work [9], authors suggested a view-based approach to perceptually characterize 3DGCs based on saliency dispersion. They extended the concept of Visual Attention Complexity (VAC) indicator, already introduced in 2D imaging context [10], by adapting it to 3DGCs. Such approach offers the possibility to associate a VAC score to each view of a 3D object.

Considering the above-presented works, we believe that visual saliency has a crucial influence on the perceived quality and that combining this information appropriately with geometric and appearance attributes can improve the perceptual quality prediction. Therefore, we propose an objective quality metric for colored 3D meshes, which is built as an extension of CMDM; state-of-the-art metric for such data. It combines the geometry and color features of CMDM, with a new perceptual feature motivated by visual saliency. The proposed metric is detailed in section 2, while its results are presented in section 3. This work attempts to make a first step towards integrating a visual saliency measure in a perceptual quality metric for 3D contents.

2. INTEGRATING VISUAL SALIENCY INTO AN OBJECTIVE QUALITY METRIC FOR COLORED MESHES

The metric we propose is an extension of the Color Mesh Distortion Measure (CMDM)[4], which is a full-reference metric that operates entirely on the mesh domain and incorporates

This work was funded by the French National Research Agency as part of ANR-PISCO project (ANR-17-CE33-0005).

accurate geometry and color features. The novelty of our metric is the integration of the Visual Attention Complexity (VAC) indicator [9] to these latter features. We refer to our metric as CMDM-VAC. An overview of the proposed metric is shown in Figure 1.

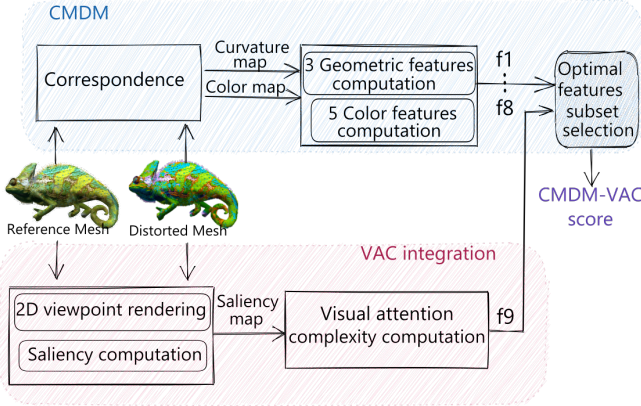


Fig. 1. Overview of the proposed metric.

2.1. CMDM concept

CMDM is based on the following steps [4]: First, a correspondence is computed between the distorted M_{dist} and the reference M_{ref} models. Then, a set of geometric and color features is computed locally, i.e., over the points belonging to a spherical neighborhood around each vertex v of M_{dist} and their corresponding points \hat{v} on M_{ref} . The neighborhood radius is set as in [4].

- **Geometry-based features:** These features are based on the curvature information [11].

$$\text{Curvature comparison } f_1(v) = \frac{|\overline{C}_v - \overline{C}_{\hat{v}}|}{\max(\overline{C}_v, \overline{C}_{\hat{v}}) + k} \quad (1)$$

$$\text{Curvature contrast } f_2(v) = \frac{|\sigma_{C_v} - \sigma_{C_{\hat{v}}}|}{\max(\sigma_{C_v}, \sigma_{C_{\hat{v}}}) + k} \quad (2)$$

$$\text{Curvature structure } f_3(v) = \frac{|\sigma_{C_v} \sigma_{C_{\hat{v}}} - \sigma_{C_v C_{\hat{v}}}|}{\sigma_{C_v} \sigma_{C_{\hat{v}}} + k} \quad (3)$$

where k is a constant set to 1 to avoid instability. $(\overline{C}_v, \overline{C}_{\hat{v}})$, $(\sigma_{C_v}, \sigma_{C_{\hat{v}}})$ and $\sigma_{C_v C_{\hat{v}}}$ are respectively Gaussian-weighted averages, standard deviations and covariance of curvature over the points belonging to the neighborhood of v and over their projection on M_{ref} .

- **Color-based features:** As explained in [12], these features are computed in the perceptually uniform color space LAB2000HL [13], in which each vertex v has of a lightness and two chromatic values (L_v, a_v, b_v) . The chroma of v is defined as: $Ch_v = \sqrt{a_v^2 + b_v^2}$.

$$\text{Lightness comparison } f_4(v) = \frac{1}{c_1(\overline{L}_v - \overline{L}_{\hat{v}})^2 + 1} \quad (4)$$

$$\text{Lightness contrast } f_5(v) = \frac{\sigma_{L_v} \sigma_{L_{\hat{v}}} + c_2}{\sigma_{L_v}^2 + \sigma_{L_{\hat{v}}}^2 + c_2} \quad (5)$$

$$\text{Lightness structure } f_6(v) = \frac{\sigma_{L_v L_{\hat{v}}} + c_3}{\sigma_{L_v} \sigma_{L_{\hat{v}}} + c_3} \quad (6)$$

$$\text{Chroma comparison } f_7(v) = \frac{1}{c_4(\overline{Ch}_v - \overline{Ch}_{\hat{v}})^2 + 1} \quad (7)$$

$$\text{Hue comparison } f_8(v) = \frac{1}{c_5 \Delta H_{v\hat{v}}^2 + 1} \quad (8)$$

with $\Delta H_{v\hat{v}} = \sqrt{(a_v - a_{\hat{v}})^2 + (b_v - b_{\hat{v}})^2 - (Ch_v - Ch_{\hat{v}})^2}$. $\Delta H_{v\hat{v}}$ refers to the Gaussian-weighted average hue difference between the neighborhood of v and its projection. The constants values c_i were set as in [12]. Note that, the color-based features are inverted ($f_j(v) = 1 - f_j(v)$) so that a value of 0 indicates that there is no local geometric and color distortion around v . All features $\in [0, 1]$.

The geometric and color features are computed for each vertex of the M_{dist} . In order to obtain global features f_j , local values of each feature are averaged over all the vertices. The global distortion (GD) score is thus defined as a linear combination of f_j , optimized through logistic regression.

$$GD_{M_{dist} \rightarrow M_{ref}} = \sum_{j \in S} w_j f_j \quad (9)$$

S is the set of feature indexes of the linear model. w_j weights the contribution of each feature to the overall distortion score. In order to strengthen the robustness of the metric and to obtain a symmetric measure, $GD_{M_{ref} \rightarrow M_{dist}}$ is computed, and the average is retained as the final distortion measure. As in [14], the optimal subset of features and their corresponding weights w_j are determined through an optimization computed through cross-validation.

2.2. Integration of the VAC measure

We highlight that the saliency information has been exploited on conventional media (e.g., images and videos) using saliency maps directly without any further processing to improve the performance of quality metrics [10]. In this work, we consider colored 3D graphical contents. Moreover, we do not use saliency maps directly but instead compute the Visual Attention Complexity (VAC) indicator adapted to 3D objects [9]. This indicator is incorporated as a perceptual feature in CMDM in order to improve its performance. The VAC indicator consists in evaluating the complexity of the visual saliency as shown in Figure 2. We distinguish views with low VAC scores indicating focused 3D contents and those with high VAC scores indicating exploratory 3D contents, as detailed later in this section. The VAC is computed as follows: Once the views of the 3D object are generated, a computational saliency model is used to compute saliency maps. We use the Salicon [15] computational model as recommended in [9]. We apply a mask on the visible surface of the 3D object with a border enlargement (using a morphological operation with a disk diameter equal to 1° of visual angle). This enlargement takes into consideration the gazing

uncertainty. Since the saliency map represents the probability of gazing at a given pixel, a normalization is applied to the masked saliency data so that the pixel values $\in [0,1]$. Finally, we compute a conditional entropy on the normalized saliency information, i.e., zero probability pixels are not included. Thus, the VAC score of a rendered 3D object view is defined as follows:

$$VAC\ score = - \sum_{i=1}^n p_i \log_2 p_i \quad (10)$$

with $p_i = h_i/K$, where h_i is the histogram for the intensity value i in the masked saliency map S (i.e., the visible surface of the 3D object), and K is the total number of pixels in S . We obtain a score associated with each view of the 3D object. Low VAC score values indicate that there are strongly salient regions on the visible surface of the rendered 3D object. This is referred to as focused views (see Figure 2.a). On the contrary, when the VAC scores are high, the saliency is diffused (i.e., overall gazing behavior). We then refer to exploratory views (see Figure 2.b).



(a) VAC score: 5.81

(b) VAC score: 7.51

Fig. 2. Saliency data and corresponding VAC scores. In (a) and (b), the rendered view of a 3D object is on the right, and its corresponding masked saliency information is on the left.

To ensure consistency with geometry-based and color-based features, we normalized the VAC feature as follows:

$$VAC\ comparison\ f_9 = \frac{|VAC_{dist} - VAC_{ref}|}{VAC_{ref}} \quad (11)$$

All features are then within the range $[0, 1]$. The VAC feature was included in the global distortion score as the 9th feature, noted as f_9 in the linear combination detailed in Eq. 9.

3. RESULTS AND EVALUATION

3.1. Database

We trained and tested our metric on a very recent database produced from a subjective study, based on the Double Stimulus Impairment Scale (DSIS) method and conducted in a virtual reality environment [4]. This database is composed of 480 dynamic meshes with vertex colors. It is the first public¹ dataset for such data. The stimuli were generated from 5 source models (“Aix”, “Ari”, “Chameleon”, “Fish”, “Samurai”) subjected to 4 types of distortion, each applied with 4 strengths. The selected distortions are uniform quantizations applied on either (1) geometry (QGeo) or (2) color (QCol), simplification algorithms that take into account either

(3) the geometry only (SGeo) or (4) both geometry and color (SCol). Each stimulus was displayed in 3 viewpoints/views and animated with two short movements. In this work, we do not consider the influence of either viewpoints or animations. Thus, for a given stimulus, we averaged its mean opinion scores (MOSs) over the different viewpoints and animations. The database used is therefore composed of 80 stimuli. We note that the VAC feature used in the following analyses is the average of VAC scores computed for each viewpoint.

3.2. Toward an optimal combination of features

Since the number of features we implemented is high, we designed two Leave-One Out Cross-Validation tests (LOOCV) in order to select the optimal subset of features, as in [4]. In fact, training the metric using all the features increases the risk of over-fitting. Moreover, features may have redundancy between each other. LOOCV tests consist of splitting the dataset into 2 parts: a training set used to optimize feature weights by logistic regression and a test set used to evaluate the obtained metric. For the first LOOCV test, the splitting is made according to the source models. This amounts to selecting 4 source models among 5 for training and using the remaining model and its distortions for testing. After 5 folds, each source model has been used as a test set. In the second LOOCV test, we divided the database according to distortion types. We trained the metric on 3 distortion types out of 4. We repeated this operation 4 times so that each distortion type is used once for testing.

Since we have 9 different features, there are 512 possible combinations of features. Hence, we exhaustively searched through all possible combinations and selected the one that generates the best average performance of our metric over all the test sets in terms of the mean of Pearson (PLCC) and Spearman (SROCC) correlations. In the end, we found that the best model of CMDM-VAC is composed of only 5 features: Curvature contrast (f_2), Lightness contrast (f_5) and structure (f_6), hue comparison (f_8) and VAC comparison (f_9). Interestingly, the visual attention complexity feature was selected among the optimal subset of features indicating the importance of the visual saliency on the perceived quality of artifacts. This is consistent with the single feature performance analysis: the (PLCC, SROCC) of f_9 with the MOSs are (0.61, 0.7). The results of the other features are reported in [4]. Note that, the selected features for CMDM are f_2, f_5, f_6 and f_7 [4].

3.3. Performance of our metric and comparisons

In figure 3, we report the average performance of our metric *CMDM-VAC* over the cross-validation test sets. For comparison purposes, we included the results of *CMDM* and 3 state-of-the-art Image Quality Metrics (IQMs): *SSIM* [16], *HDR-VDP2* [17], *iCID* [18], reported in [4].

In both LOOCV tests, *CMDM-VAC* performs better than *CMDM*. Considering the LOOCV test among the distortions,

¹<https://yananehme.github.io/datasets/>

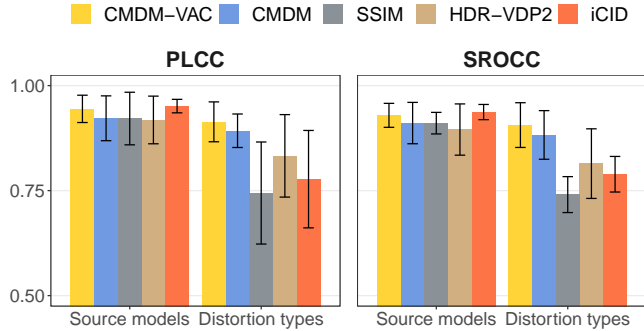


Fig. 3. Pearson (PLCC) and Spearman (SROCC) correlations of several metrics with MOSs of the used database [4]. Mean performance evaluation measures are reported. Error bars indicate the standard deviation over the test sets.

CMDM-VAC and *CMDM* outperform IQMs showing that the latter have difficulties in ranking distortions applied on different sources [1].

The most noticeable improvements of *CMDM-VAC* were observed for the 2 test sets presented in Table 1. As reported in [4], *CMDM* exhibits a poor performance when assessing the quality of stimuli geometrically quantized (QGeo) due to the fact that this distortion superimposes the vertices of the stimulus, which leads to not knowing the exact vertex color taken into account in the rendering pipeline. Table 1 shows that integrating the VAC considerably improves the results. Indeed, this measure is computed on a snapshot of the stimulus that naturally incorporates the entire rendering conditions.

Table 1. Performance comparison of several metrics on some test sets of the 2 LOOCV tests. For metrics marked with a *, the values are reprinted from [4].

	QGeo		Chameleon	
	PLCC	SROCC	PLCC	SROCC
CMDM-VAC	0.926	0.864	0.89	0.89
CMDM*	0.882	0.825	0.83	0.83
SSIM*	0.875	0.794	0.823	0.868
HDR-VDP2*	0.946	0.938	0.849	0.818
iCID*	0.88	0.838	0.924	0.921

Regarding the second test set, it denoted in [4] that of all the source models, the metrics perform less well on the model having the most content information/details: the Chameleon. Table 1 shows that saliency is potentially an important cue for a more effective quality assessment of complex and rich models.

3.4. Recommended weights

To provide the final model of *CMDM-VAC*, we averaged the weights obtained for each of the 9 training sets from the 2 LOOCV tests. The recommended weights, as well as the importance of each feature, are reported in Table 2.

The (PLCC, SROCC) computed over the whole dataset (80 stimuli) for *CMDM-VAC* and *CMDM* are (0.936, 0.922)

Table 2. Weights and importance of the selected features. The importance values are defined as the weights scaled with the standard deviation of the features.

Features	f_2	f_5	f_6	f_8	f_9
Recommended weights	0.092	0.202	0.028	0.182	0.496
Importance	0.293	0.359	0.051	0.165	0.131

and (0.913, 0.9), respectively. The improvement in our metric results, after integrating the VAC, is statistically significant (with a p-value=0.0094, obtained by a statistical test on the logistic regression residual of the 2 metrics).

3.5. Performance evaluation per quality range

To assess whether our metric is vulnerable to the quality range of stimuli, we divided the database into 2 groups: (1) low-quality stimuli having MOSs $\in [1,3[$ and (2) good quality stimuli having MOSs $\in [3,5]$. Results are shown in Figure

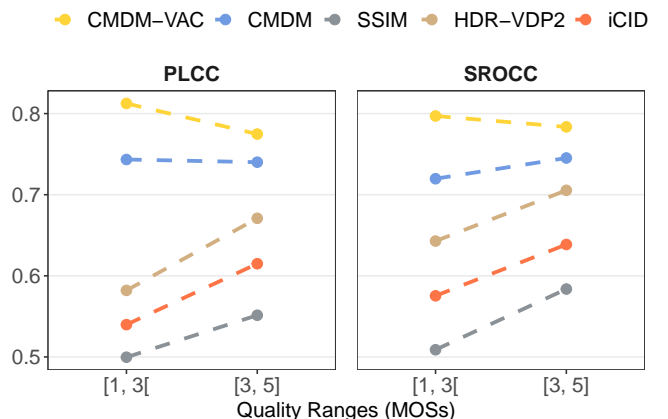


Fig. 4. Performance evaluation of several metrics according to the quality range of stimuli.

Among the IQMs, *HDR-VDP2* seems the best choice for estimating the perceived quality, especially for near-threshold distortions (MOSs > 3). Unlike IQMs, *CMDM-VAC* performs slightly better on low-quality stimuli than on higher quality ones. Moreover, the overall performance of our metric seems to be more stable over quality ranges than that of IQMs.

4. CONCLUSION

In this work, we extended the perceptual quality metric *CMDM*, designed for 3D meshes with diffuse colors: we combined its geometry and color features with a perceptual feature motivated by visual saliency.

Including the visual attention complexity (VAC) feature seems promising to improve the visual quality prediction.

In fact, integrating this feature in *CMDM* showed an overall improvement of its performance, especially when assessing the quality of stimuli geometrically quantized. As future work, it would be interesting to explore how the VAC can improve other quality metrics.

5. REFERENCES

- [1] G. Lavoué, M. C. Larabi, and L. Vasa, “On the Efficiency of Image Metrics for Evaluating the Visual Quality of 3D Models,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 8, pp. 1987–1999, 2016.
- [2] D. Tian and G. AlRegib, “Batex3: Bit allocation for progressive transmission of textured 3-d models,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 23–35, 2008.
- [3] J. Guo, V. Vidal, I. Cheng, A. Basu, A. Baskurt, and G. Lavoué, “Subjective and objective visual quality assessment of textured 3D meshes,” *ACM Transactions on Applied Perception*, vol. 14, no. 2, 2016.
- [4] Y. Nehmé, F. Dupont, J. P. Farrugia, P. Le Callet, and G. Lavoué, “Visual quality of 3d meshes with diffuse colors in virtual reality: Subjective and objective evaluation,” *IEEE Transactions on Visualization and Computer Graphics*, , no. 1, pp. 1–18, 2020.
- [5] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- [6] G. Leifman, E. Shtrom, and A. Tal, “Surface regions of interest for viewpoint selection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 12, pp. 2544–2556, 2016.
- [7] Y. Zhao, Y. Liu, R. Song, and M. Zhang, “A saliency detection based method for 3d surface simplification,” in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 889–892.
- [8] X. Jiao, T. Wu, and Xuzhou Qin, “Mesh segmentation by combining mesh saliency with spectral clustering,” *J. Comput. Appl. Math.*, vol. 329, pp. 134–146, 2018.
- [9] M. Abid, M. Perreira Da Silva, and P. Le Callet, “Perceptual characterization of 3d graphical contents based on attention complexity measures,” in *QoEVMA’20: Proceedings of the 1st Workshop on Quality of Experience (QoE) in Visual Multimedia Applications*, New York, NY, USA, 2020, p. 31–36.
- [10] W. Zhang, R. R. Martin, and H. Liu, “A saliency dispersion measure for improving saliency-based image quality metrics,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 6, pp. 1462–1466, 2018.
- [11] P. Alliez, D. Cohen-Steiner, O. Devillers, B. Lévy, and M. Desbrun, “Anisotropic polygonal remeshing,” *ACM Trans. Graph.*, vol. 22, no. 3, pp. 485–493, July 2003.
- [12] I. Lissner, J. Preiss, P. Urban, M. S. Lichtenauer, and P. Zolliker, “Image-difference prediction: From grayscale to color,” *IEEE Transactions on Image Processing*, vol. 22, no. 2, pp. 435–446, 2013.
- [13] I. Lissner and P. Urban, “Toward a unified color space for perception-based image processing,” *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 1153–1168, 2012.
- [14] Y. Liu, J. Wang, S. Cho, A. Finkelstein, and S. Rusinkiewicz, “A no-reference metric for evaluating the quality of motion deblurring.,” *ACM Transactions on Graphics*, vol. 32, no. 6, pp. 171–175, 2013.
- [15] M. Jiang, S. Huang, J. Duan, and Q. Zhao, “Salicon: Saliency in context,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1072–1080.
- [16] Z. Wang, A. C. Bovik, H.R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [17] R. Mantiuk, K. J. Kim, A. G. Rempel, and W. Heidrich, “Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions,” *ACM Trans. Graph.*, vol. 30, no. 4, July 2011.
- [18] J. Preiss, F. Fernandes, and P. Urban, “Color-image quality assessment: From prediction to optimization,” *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1366–1378, 2014.